

(11)Publication number : 2000-305578
(43)Date of publication of application : 02.11.2000

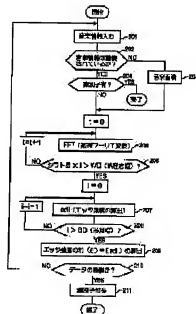
G10L 11/00
G06F 12/00
G06F 17/30
G10L 15/02
G10L 21/06

(71)Applicant : NIPPON TELEG & TELEPH CORP
<NTT>

(72)Inventor : MINAMI KENICHI
MATSUURA YUMIKO
TOMIOKA ATSUKI
KOJIMA HARUHIKO

PROBLEM TO BE SOLVED: To add an identifier to music information in real time with a relatively simple processing.

SOLUTION: A spectrogram is calculated by analyzing frequency of input music information 205. Edge intensity (edi) of the spectrogram in the time axis direction is calculated 207. The edge intensity (edi) is summed in the frequency axis direction to calculate an edge intensity sum ED 209. Data indicating the ED or an increase, a decrease or a continuation of the ED are added to a header of the music information as an identifier to store in a database.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-305578

(P2000-305578A)

(43) 公開日 平成12年11月2日 (2000.11.2)

(51) Int.Cl. ⁷	識別記号	F I	テームド (参考)
G 1 0 L 11/00		G 1 0 L 7/02	A 5 B 0 7 5
G 0 6 F 12/00	5 4 7	G 0 6 F 12/00	5 4 7 D 5 B 0 8 2
		15/40	3 7 0 E 5 D 0 1 5
G 1 0 L 15/02		15/401	3 1 0 C
21/06			3 3 0 Z

審査請求 未請求 請求項の数 9 O L (全 6 頁) 最終頁に続く

(21) 出願番号 特願平11-118208

(22) 出願日 平成11年4月26日 (1999.4.26)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72) 発明者 南 薫一

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72) 発明者 松浦 由美子

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74) 代理人 100066153

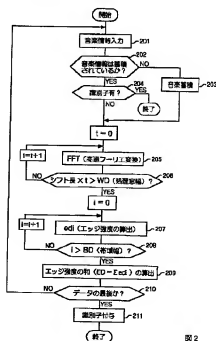
弁理士 草野 卓 (外1名)

最終頁に続く

(54) 【発明の名称】 音楽データベース作成装置、作成方法およびそのプログラム記録媒体

(57) 【要約】

【課題】 比較的単純な処理により実時間で音楽情報に識別子を付与できる。

【解決手段】 入力音楽情報を周波数分析してスペクトログラムを算出し (205)、その時間軸方向のエッジ強度 $ed1$ を算出し (207)、エッジ強度 $ed1$ を周波数方向において加算してエッジ強度 E と D を算出し (209)、その E D 又は E D が増加、減少か、継続かを示すデータを識別子として音楽情報のヘッダに付けてデータベースに格納する。

【特許請求の範囲】

【請求項1】 音楽情報を示す音響信号を入力し各々区別して記録手段に記録する音楽データベース作成装置において、

前記音響信号を周波数分析して周波数スペクトルを算出する手段と、

前記周波数スペクトルのピークの周波数方向に対する安定性を示す安定性係数を算出する手段と、

前記安定性係数の時間的変化に基づいた識別子を決定する手段とを有し、

前記記録手段は前記識別子と前記音響信号とを対応付けて記憶することを特徴とする音楽データベース作成装置。

【請求項2】 前記安定性係数算出手段は、前記周波数分析結果をスペクトログラムとして画像情報に変換する手段と、

前記スペクトログラムから時間方向のエッジ強度を算出し、

前記エッジ強度を安定性係数とする手段とよりなることを特徴とする請求項1記載の音楽データベース作成装置。

【請求項3】 前記識別子決定手段は、前記エッジ強度の時間的変化を、増加、減少、継続、の3パターンのうち何れかで表現し、

前記パターンの時間的変化を識別子とする手段であることを特徴とする請求項2記載の音楽データベース作成装置。

【請求項4】 音楽情報を示す音響信号を入力し各々区別して記録手段に記録する音楽データベース作成方法において、

前記音響信号を周波数分析して周波数スペクトルを算出する過程と、

前記周波数スペクトルのピークの周波数方向に対する安定性を示す安定性係数を算出する過程と、

前記安定性係数の時間的変化に基づいた識別子を決定する過程とを有し、

前記記録手段は前記識別子と前記音響信号とを対応付けて記憶することを特徴とする音楽データベース作成方法。

【請求項5】 前記安定性係数算出過程は、前記周波数分析結果をスペクトログラムとして画像情報に変換する過程と、

前記スペクトログラムから時間方向のエッジ強度を算出し、

前記エッジ強度を安定性係数とする過程とよりなることを特徴とする請求項4記載の音楽データベース作成方法。

【請求項6】 前記識別子決定過程は、前記エッジ強度の時間的変化を、増加、減少、継続、の3パターンのうち何れかで表現し、

前記パターンの時間的変化を識別子とすることを特徴とする請求項5記載の音楽データベース作成方法。

【請求項7】 前記音響信号を周波数分析して周波数スペクトルを算出する処理と、

前記周波数スペクトルのピークの周波数方向に対する安定性を示す安定性係数を算出する処理と、

前記安定性係数の時間的変化に基づいた識別子を決定する処理と、

前記識別子と前記音響信号とを対応付けて記録手段に記憶する処理と、をコンピュータが実行することを特徴とする音楽データベース作成処理プログラム記録媒体。

【請求項8】 前記安定性係数算出処理は、前記周波数分析結果をスペクトログラムとして画像情報に変換する処理と、

前記スペクトログラムから時間方向のエッジ強度を算出し、

前記エッジ強度を安定性係数とする処理とよりなることを特徴とする請求項7記載の音楽データベース作成処理プログラム記録媒体。

【請求項9】 前記識別子決定処理は、前記エッジ強度の時間的変化を、増加、減少、継続、の3パターンのうち何れかで表現し、

前記パターンの時間的変化を識別子とすることを特徴とする請求項8記載の音楽データベース作成処理プログラム記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音楽信号から曲ごとに異なる特徴量を算出し、特徴量を識別子として音楽信号に対応づけてデータベースに記憶する方法、装置およびその方法を実現するためのプログラムを記録した記録媒体に関する。

【0002】

【従来の技術】音楽情報にID（識別子）を付与する方法としては、ピッチの変化パターンを求める方法（陸山他、「ハミング歌唱を手掛かりとするメロディ検索」、電子情報通信学校論文誌、Vol. 1、777-D-II、No. 8、1994）や特徴ベクトルのヒストグラムに基づく方法（柏野他、「マルチモーダルアクティブ探索を用いた画像・音響時系列の高速探索」、電子情報通信学会技術報告、PRMU98-80、1998）、MIDI信号のような中間記述に変換する方法、また近年では電子透かしによる方法などが提案されている。これらの技術は、音楽の検索、管理、著作権保護などに利用されている。

【0003】また、音楽信号のスペクトルの安定性を求める方法は、様々な音源が混在している音信号から音楽を検出する方法（南他、「音解析による映像インデクシング」、電子情報通信学会総大会、D-1-2-64、1997）として提案されている。

【0004】

【発明が解決しようとする課題】ピッチの変化パターンを求める方法は、ピッチの抽出を正確に行うことが難しいため、童謡などの単純な音楽にしか対応できない。ヒストグラムを利用する方法では、既知の音楽に対してヒストグラムを作成しておかなければならない。例えば放送音声で未知の音楽が流れた場合には、音楽であることを人間が判断し、音楽の区間を切り出してヒストグラムを求め、それを登録する、といった作業が必要になる。未知の音楽に対して実時間で処理が困難である。MIDI信号に変換する方法では、予めメロディが分かっているといないと交換が難しく、電子透かしでは音質の劣化が問題となる。

【0005】このように、従来の方法はある条件下においては有効であるものの、実時間性と簡便性を備えたものはなかった。一方、スペクトルの安定性を求める方法では、音声信号と混った音楽の検出を行うことが可能になる。未知の音楽が存在した場合でもその区間を特定できる。しかし、音楽データの識別を扱ったものではない。

【0006】この発明は、比較的単純な処理により、実時間かつ少ない情報量で音楽信号に識別子を付与して記憶することができ音楽データベース作成方法、装置およびプログラム記録媒体を提供することを目的とする。

【0007】

【課題を解決するための手段】この発明によれば音楽信号を周波数分析して、周波数スペクトルを求め、その周波数スペクトルのピークの周波数方向に対する安定性を示す安定性係数を算出し、その安定性係数の時間的変化に基づいた識別子を決定し、その識別子と音楽信号とを対応付けて記録手段に記憶する。

【0008】安定性係数の算出は、周波数分析結果をスペクトログラムとして画像情報に変換し、そのスペクトログラムから時間方向におけるエッジ強度を算出し、そのエッジ強度を安定性係数とする。更に識別子の決定は、エッジ強度の時間的変化を、増加、減少、継続の3パターンのうち何れかで表現し、そのパターンの時間的変化を識別子とする。

【0009】この発明では周波数方向に対する安定性を特徴量として用いる。「周波数方向に対する安定性」は、スペクトルのピークの周波数方向の（一定時間当たりの）変動量の逆数、即ちエッジ強度の和として算出される。音楽では特定のピッチ成分が支配的であるため、その周囲の周波数成分よりも優勢となり周波数方向にわたる差分（微分）が大きくなる。そのためにエッジ強度の和の値が高くなると音楽であるか否かが判断できる。

【0010】ここで、エッジ強度の和が一定時間毎に求められ、その時系列を識別子として求める。この識別子は音楽によって異なるので、音楽区間を切り出すための手がかりとなる。従って、識別子と音楽データとを対応

付けてデータベースとして記憶できれば、所望の音楽データを実時間で認識できるようになる。一定時間（処理フレーム）毎に算出されるエッジ強度の時間的変化を増加・減少・継続の3パターンで表現することにより少ない情報量で識別子を付与することができる。例えば、エッジ強度が時間の経過と共に、{50, 20, 30, 30, 60, 100, 120, 100}と求まった時に、増加を01、減少を10、継続を00と表せば{10, 01, 00, 01, 01, 01, 10}というように識別子を決定できる。

【0011】識別子の要素となるエッジ強度をそのままの数値で表現すると、1つの値に8bit必要であるが、上記3パターンで表現すれば、1つの値に2bitで十分である。そのため、識別子の情報量を少なくすることができる。検索においては、この識別子を手がかりとして用いることができるようになる。ここで、例えば、DPマッチング、エッジ強度分布の類似性を手がかりとして用いられよう。

【0012】

20 【発明の実施の形態】以下に、この発明の実施例について図面を参照して説明する。図1は、この発明の一実施形態の音楽データベース作成装置の概略構成を示すブロック図である。この実施形態の音楽データベース作成装置は、音楽情報を入力する音楽入力部101と、音楽情報を蓄積する音楽蓄積部102と、音楽情報の信号を周波数解析する周波数解析部103と、周波数スペクトルのピークの周波数方向に対する安定性を算出する安定性算出部104と、安定性の時間的変化を識別子として音楽情報に付与する識別子付与部105から構成されている。

30 【0013】データベース音楽情報と識別子とを同一のファイルで管理する場合、入力音楽情報に識別子を付与した後で、蓄積する。また、別個のファイルで管理する場合も考えられる。そのとき、入力音楽情報に識別子を付与した後で、入力音楽情報が予め蓄積されている識別子を蓄積し、入力音楽情報が予め蓄積されていない音楽情報および識別子を蓄積する。

40 【0014】データベースに識別子と音楽情報とを関連づけて記憶するには、音楽ファイルと識別子を別個のファイルとして管理する場合、音楽のファイル名のテーブルと識別子のテーブルに連番を付与する。識別子と音楽情報とを同一のファイルで管理する場合には、音楽ファイルのヘッダ情報として識別子を記憶させる、または音楽情報に識別子を電子透かしとして埋め込むことが考えられる。

50 【0015】図2は、この発明の一実施例の音楽識別装置の処理の流れを示したフローチャートである。この発明をソフトウェアで実現した場合でも同様の処理の流れとなる。まず、音楽情報を入力部101で入力し、判断処理202において、その入力された音楽情

報が音楽蓄積部 102 に予め蓄積されているデータかどうかを判断し、予め蓄積されていなければ、音楽蓄積処理 203 でその音楽情報を音楽蓄積部 102 に蓄積する。音楽情報が音楽蓄積部 102 に予め蓄積されていれば、処理 204 でその音楽情報に識別が付与されているかを調べ、付与されていれば処理を終了し、識別子が付与されている場合、及び音楽蓄積部 102 に蓄積されていない場合は、次に F F T 処理 205 において入力された音楽情報の信号を適当なフレーム長とフレームシフトで F F T (高速フーリエ変換) 処理する。フレーム長とフレームシフトは、サンプリング周波数が 8 k H z のときに各々 512 ポイント、80 ポイント程度が良い。サンプリング周波数が高い場合には、信号にローパスフィルタをかけたのちダウンサンプリングすることも可能である。F F T 処理はフレームシフトの回数を t としたときにフレームシフト長 $\times t$ が処理窓幅 $W D$ になるまで繰り返す。その判断を判断処理 206 において行う。処理窓幅 $W D$ は通常 0.5 s 程度の長さで行うが、任意に設定することも可能である。尚、F F T 処理以外に、ウェーブレット変換やフィルタバンク等の周波数解析手法を用いることも可能である。

【0016】F F T 処理が終了すると、処理窓幅 $W D$ の時間幅を持つサウンドスペクトログラムが得られる。図 3 A は、10 s 程度のサウンドスペクトログラムの例を示している。このような画像に対し、周波数 i における時間方向のエッジ強度 $e d i$ をエッジ強度算出処理 207 で求める。エッジ強度は周波数方向の差分値を時間方向に足し合わせるにより求めることが可能である。つまりスペクトログラムを、 x 方向を j (時間軸)、 y 方向を i (周波数軸) とした画像として考えて、周波数 i におけるエッジの強さ $e d i$ を次式で算出する。

$$\text{【0017】 } e d i = \sum_{j=0}^{N-1} \{ f(j, i-1) - f(j, i) + f(j, i+1) \}$$

上式はデジタル画像についてエッジを 2 次差分 (アナログ画像では 2 次微分) として求めた場合の例である。ここで、 $f(j, i)$ は画像 (j, i) の輝度、 t はエッジを検出する時間方向の範囲の長さ = 処理窓幅 $W D$ で、 $i = \{1, 2, 3, \dots, N-1\}$ (N は周波数方向の画素数 = 帯域幅 $B D$) である。

【0018】エッジの強さを 1 次差分 (微分) から定義することも可能である。その場合次のように計算される。

$$e d i = \sum_{j=0}^{N-1} \{ f(j, i+1) - f(j, i) \}$$

この処理を周波数 i が帯域幅 $B D$ になるまで繰り返す。この繰り返しの判断を処理 209 で行う。処理 209 で $e d i$ を帯域幅 $B D$ 分加算して $E D = \sum_{i=1}^{B D} |e d$

$i|$ を求めてエッジ強度の和 $E D$ を算出する。

【0019】以上の処理を信号全体に対して行った場合の一部の区間におけるエッジ強度 $E D$ を図 3 B に示す。信号全体に対して以上の処理を行ったかの判断を処理 210 で行う。識別子付与処理 211 では、蓄積された音楽情報に、求められた $E D$ を識別子として書き込む。書き込む場所は、音楽情報のヘッダ部分でも良いし、音楽情報を適当な長さを持つ区間に分割し、区間ごとに書き込んで良い。また、エッジ強度 $E D$ のものの値を書き込む他、エッジ強度 $E D$ の値が増加したか減少したかあるいは変化がないか、といった情報だけを書き込むことも可能である。さらに、識別子に割り当てられるビット長に応じて、適当な時間間隔で $E D$ の平均値や最大値、あるいは $E D$ があるしきい値以上かかといった情報などを識別子とすることも可能である。

【0020】予めデータベース (音楽蓄積部) に記憶されたものに識別子を付与して再びデータベースに記憶する場合もある。

【0021】

【発明の効果】(1) 請求項 1、4 および 7 の発明によれば、音楽情報の信号を周波数解析し、周波数スペクトルのピークの周波数方向に対する安定性を算出し、前記安定性の時間的変化を識別子として音楽情報に付与することによって、音楽情報の識別が可能であるから、実時間で効率的に音楽データベースを作成することができる。

【0022】(2) 請求項 2、5 および 8 の発明は、周波数スペクトルのピークの周波数方向に対する安定性を算出する際に、周波数解析によって算出されたサウンドスペクトログラムから時間軸方向のエッジ強度を算出し、エッジ強度を安定性の指標として用いることにより、曲ごとに異なる特徴量 (識別子) を容易に求めることが可能となる。

【0023】(3) 請求項 3、6 および 9 の発明は、エッジ強度の時間的変化を増加、減少、継続の 3 パターンの何れかで表現することにより、少ない情報量で音楽を識別することが可能となる。

【図面の簡単な説明】

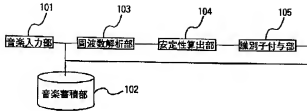
【図 1】この発明の一実施形態の音楽データベース作成装置の概略機能構成を示すブロック図。

【図 2】この発明の一実施形態の音楽データベース作成装置の処理の流れとこの発明をソフトウェアで実現した場合の処理の流れを示すフローチャート。

【図 3】この発明の一実施形態の音楽データベース作成装置において得られるサウンドスペクトログラムとエッジ強度の様子を示す図。

【図1】

図1



【図2】

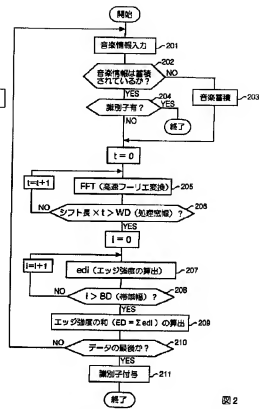


図2

【図3】

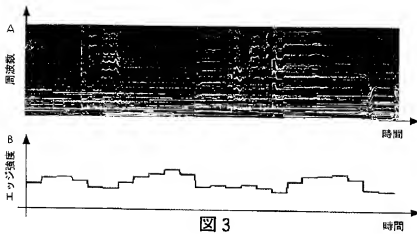


図3

フロントページの続き

(51)Int.Cl.

識別記号

F I

ターマコード (参考)

G I O L 3/00

5 1 5 A

7/10

(72)発明者 富岡 淳樹

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72)発明者 児島 治彦

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

F ターム(参考) 5B075 ND14 NK06 NK13 NR05

5B082 AA00 EA01 EA37 CA07

5D015 AA06 FF01